

University of Groningen

OC-0418: Quantitative evaluation of deep learning contouring of head and neck organs at risk

Peressutti, D.; Aljabar, P.; Dijk, L.V. Van; Bosch, L. Van den; Gooding, M.; Brouwer, C.L.

DOI:

[10.1016/S0167-8140\(18\)30728-X](https://doi.org/10.1016/S0167-8140(18)30728-X)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Peressutti, D., Aljabar, P., Dijk, L. V. V., Bosch, L. V. D., Gooding, M., & Brouwer, C. L. (2018). *OC-0418: Quantitative evaluation of deep learning contouring of head and neck organs at risk*. S217 - S218. [https://doi.org/10.1016/S0167-8140\(18\)30728-X](https://doi.org/10.1016/S0167-8140(18)30728-X)

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

deliveries were simulated with the treatment plan subdivided into 670 fragments. For comparison 128 *rand. start* simulations with the breathing curve of the patient are shown in Fig. 2. In both cases the mean 4D-CT calculated dose (orig. plan calculated on all ten 4D-CT phases and warped to reference phase) and the mean of the random simulated doses (orange and dark blue line) almost coincide. The dose deviations for the 128 runs are very similar for both methods. (Average dose deviations for $D_{2\%}$: $\sigma=0.41\%$, $D_{50\%}$: $\sigma=0.25\%$ and $D_{98\%}$: $\sigma=0.68\%$ for both methods)

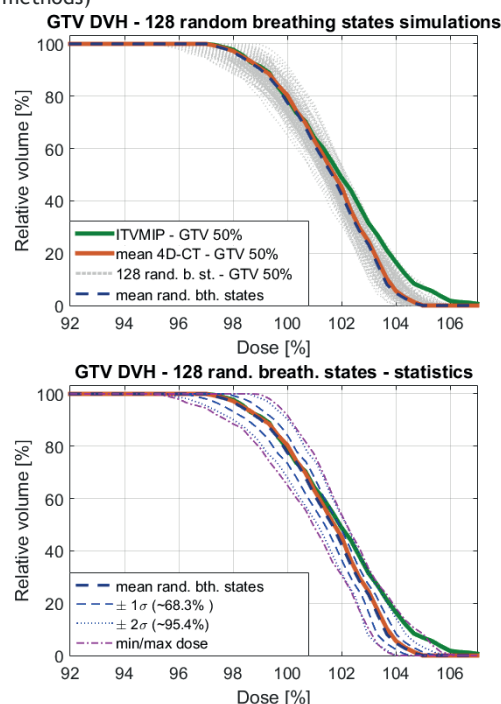


Fig. 1: DVHs showing the dose to the reference GTV (50% 4D-CT phase): orig. plan (ITVMIP), mean 4D-CT dose and 128 *rand. breath. states* simulations

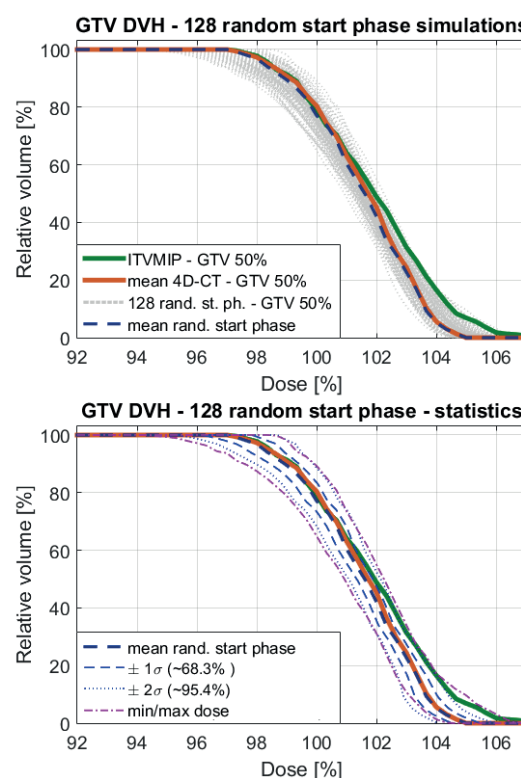


Fig. 2: Same as Fig. 1 but for 128 *rand. start* simulations

Conclusion

The *rand. breath. states* sampling is a promising method to address plan- and technique-specific interplay effects using a statistical breathing approach. Providing a patient independent statistical interplay evaluation it has the potential to comprehensively include breathing motion induced interplay effects in the pretreatment evaluation process.

Proffered Papers: RTT 4: Image acquisition and registration

OC-0418 Quantitative evaluation of deep learning contouring of head and neck organs at risk

H. Bakker¹, D. Peressutti², P. Aljabar², L.V. Van Dijk¹, L. Van den Bosch¹, M. Gooding², C.L. Brouwer¹

¹University of Groningen- University Medical Center Groningen, Department of Radiation Oncology, Groningen, The Netherlands

²Mirada Medical Ltd., Department of Radiation Oncology, Oxford, United Kingdom

Purpose or Objective

Auto-contouring has been shown to save time and improve consistency. However, despite advances in auto-contouring methods, automatically generated contours still require significant editing before they are considered clinically acceptable, in particular for structures of small size or with high anatomical variability. In this investigation, the performance of a deep learning contouring (DLC) system (WorkflowBox 2.0alpha, Mirada Medical Ltd, Oxford, UK), for the automatic contouring of organs at risk (OARs) in head and neck cancer patients has been assessed.

Material and Methods

A set of 698 head and neck patients, each comprising a CT volume image and corresponding clinical contours, was considered for this study. All cases were

acquired at a single institution. Evaluation was performed on 22 OARs in the head and neck according to international consensus delineation guidelines, comprising the arytenoids, carotid arteries, buccal mucosas, brainstem, cerebellum, cerebrum, cricopharyngeal inlet, mandible, extended oral cavity, parotid and submandibular glands, thyroid, glottic and supraglottic area, pharynx constrictor muscles, cervical esophagus and spinal cord. The set of clinical cases was randomly divided into a training set (549), cross-validation set (40) and test set (109) for training of the DLC models. Training of DLC was performed on-site. DLC was compared against an atlas-based auto-segmentation (ABAS) method (WorkflowBox 1.4) that employed a representative set of 30 atlases selected from the training set, to contour the test images. A quantitative evaluation against ground-truth clinical contours was performed by computing the Dice similarity coefficient (Dice), and average distance (AD) between both sets of automatically generated contours and the manual clinical contours.

Results

Quantitative results for the test set are shown in Figure 1 and Table 1 for the considered OARs. Figure 1 (top) shows Dice values for ABAS (x-axis) and DLC (y-axis). DLC outperforms ABAS if the symbol lies above the bisector line. Similarly, Figure 1 (bottom) shows AD values in mm for ABAS (x-axis) and DLC (y-axis). In this case, DLC outperforms ABAS if the symbol lies below the bisector line. Results from the performed evaluation show DLC to significantly outperform ABAS for 17 out of 22 OARs considered.

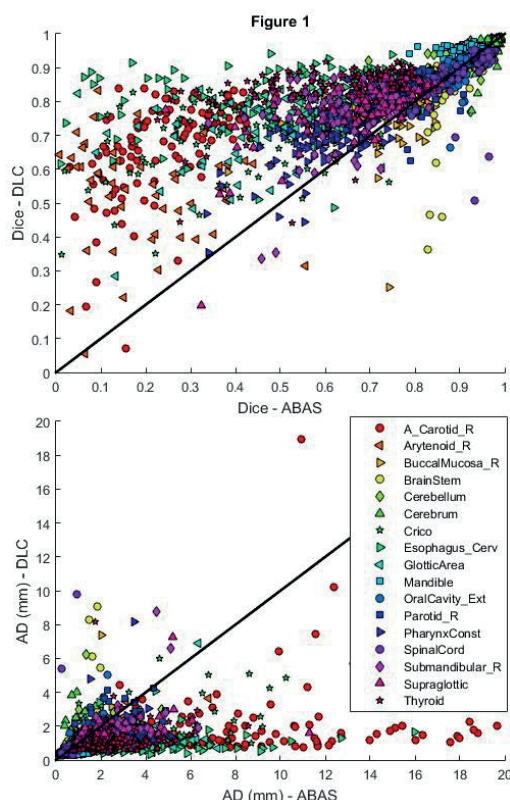


Table 1. Quantitative results of contouring accuracy of ABAS (n=109) and DLC (n=109). Median and inter-quartile ranges are shown for Dice and average contour distance. Shaded cells highlight a statistically significant better performance (Mann-Whitney U-test, $p < 0.01$).

	Dice				AD (mm)			
	ABAS		DLC		ABAS		DLC	
	Med	IQR	Med	IQR	Med	IQR	Med	IQR
Cerebrum	0.98	0.02	0.98	0.01	0.67	0.43	0.72	0.31
Cerebellum	0.94	0.03	0.94	0.03	1.14	0.60	1.13	0.47
SpinalCord	0.95	0.03	0.93	0.04	0.34	0.28	0.41	0.20
Mandible	0.93	0.04	0.95	0.02	0.49	0.32	0.36	0.11
OralCavity_Ext	0.91	0.04	0.91	0.04	1.47	0.54	1.33	0.38
BrainStem	0.87	0.04	0.87	0.04	1.35	0.54	1.25	0.45
Thyroid	0.61	0.23	0.85	0.06	2.57	1.84	0.93	0.40
Parotid_L	0.79	0.08	0.85	0.05	2.35	0.93	1.44	0.59
Parotid_R	0.78	0.09	0.83	0.06	2.44	1.31	1.46	0.58
Esophagus_Cerv	0.55	0.35	0.84	0.10	2.54	2.89	0.71	0.43
Submandibular_L	0.69	0.10	0.80	0.10	2.08	0.86	1.44	0.81
Submandibular_R	0.68	0.13	0.80	0.07	2.25	0.96	1.32	0.64
Supraglottic	0.71	0.14	0.80	0.08	1.74	1.16	1.12	0.61
BuccalMucosa_L	0.72	0.11	0.77	0.06	1.81	0.82	1.33	0.46
BuccalMucosa_R	0.75	0.10	0.77	0.07	1.60	0.82	1.35	0.51
GlotticArea	0.51	0.21	0.74	0.10	2.26	1.67	1.04	0.50
Crico	0.48	0.24	0.73	0.11	2.86	2.44	1.51	0.81
PharynxConst	0.62	0.11	0.70	0.07	1.94	0.72	1.26	0.38
A_Carotid_L	0.36	0.20	0.73	0.13	4.16	2.69	1.32	0.90
A_Carotid_R	0.25	0.17	0.72	0.14	9.16	7.69	1.51	1.57
Arytenoid_L	0.16	0.17	0.60	0.20	3.25	2.29	1.33	0.92
Arytenoid_R	0.18	0.21	0.57	0.13	3.05	1.90	1.34	0.69

Conclusion

This quantitative investigation has shown that DLC significantly outperforms ABAS methods for the automatic contouring of the majority of OARs in head and neck cancer, particularly for structures with high anatomical variability, such as parotid and submandibular glands, thyroid or small and elongated structures, such as carotids, arytenoids, glottis and supraglottic area, pharynx constrictor muscles. This improvement can be explained considering the larger amount of flexibility allowed by the deep learning models compared to ABAS. Further evaluation will aim to quantify the impact on clinical workflow and clinical outcome of the observed accuracy improvements.

OC-0419 Comparison of auto-contouring methods for regions of interest in prostate CT

P. Aljabar¹, D. Peressutti¹, E. Brunenberg², R. Smeenk², R. Van Leeuwen², M. Gooding¹

¹Mirada Medical Limited, Science Group, Oxford, United Kingdom

²Radboud University Medical Centre, Radiation Oncology, Nijmegen, The Netherlands

Purpose or Objective

Automated segmentation methods are an important part of clinical protocols for contouring regions of interest (ROIs) [1]. However, time and effort required to edit auto-contours before clinical use motivates improvements to automated methods. This study compares an established atlas-based automatic segmentation (ABAS) method against a recent deep learning contouring (DLC) approach. Both methods were used to contour ROIs in a group of prostate cancer patients.